AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 107:305–313 (1998)

The Use of the Supero-Inferior Femoral Neck Diameter as a Sex Assessor

RYAN M. SEIDEMANN,* CHRISTOPHER M. STOJANOWSKI, AND GLEN H. DORAN
Department of Anthropology, Florida State University, Tallahassee, Florida 32306

KEY WORDS femoral neck; sex assessment

ABSTRACT The present study examines the sexing potential of the minimum supero-inferior femoral neck diameter in Caucasians and African-Americans who lived at the turn of the century. A Student's *t*-test and an ANOVA indicate that population differences in neck morphology exist, albeit the strength of the test is fairly weak ($P = 0.015$). Predictive models were developed using a linear discriminant function analysis for the African-American sample, the Caucasian sample, and the combined African-American and Caucasian (AAC) sample. Jackknifed classification matrices produced classification success rates ranging from 87 to 92%. Each of the three discriminant functions were evaluated using an independent, random holdout sample. Although a smaller holdout sample usually better approximates the true error involved in an application, this was clearly not the case in this study. For African-Americans, 28 of 28 individuals were correctly classified, for Caucasians 24 of 25, and for the combined AAC sample 53 of 53 individuals were sexed correctly. It is more likely that the true accuracy of the model for the population approximates 90%. This accuracy combined with the high rate of preservation of the femoral neck makes this measurement useful in extremely fragmentary samples. Am J Phys Anthropol 107:305–313, 1998. © 1998 Wiley-Liss, Inc.

The femoral neck is a flattened expanse of bone which serves to unite the head of the femur with the shaft. The minimum diameter of the neck is found in the middle of the process with both ends expanding to form the connection with the rest of the bone. The lateral portion is of greater diameter than the medial and the neck as a whole has a slight curvature so that the anterior portion is slightly convex (Walmsley, 1915). Unlike the other anatomical landmarks of the proximal end of the femur, the neck is not the result of epiphyseal ossification. Instead it is considered to be an upward growth of the shaft and actually develops from the diaphyseal center (Paterson, 1929).

Throughout the past century and a half a considerable amount of literature has been devoted to the study of skeletal sex. Many conventional, nonmetric methods have been developed; however, there still remains a degree of ambiguity when faced with an incomplete, or even sometimes a complete, skeleton. Whereas the conventional methods typically rely on the crania and the pelvic regions, which are susceptible to considerable deterioration, this analysis focuses on an anatomical landmark which has a high rate of intact preservation. The present study is a reexamination of a univariate method for sex assessment using the minimum supero-inferior diameter of the femoral neck.

Grant sponsor: Florida State University COFRS grant system.

*Correspondence to: Ryan M. Seidemann, Department of Geography and Anthropology, Louisiana State University, Baton Rouge, LA 70803. E-mail: rseidem@unixl.sncc.lsu.edu

Received 9 January 1997; accepted 3 August 1998.

Background

Although a number of works currently exist in the literature that consider femoral neck morphology, none have directly examined the potential of this feature as a means of sex assessment. Relevant and related publications include Van Gerven's (1972) considerable work on the femur. This research, as well as the research in his dissertation (Van Gerven, 1971), is closely related to the current study. He conducted multivariate analyses of 17 measurements on the human femur, including what appears to be the exact measurement utilized in the present study. Van Gerven examined what he termed the medial-lateral diameter (MLD), and defined it as "the minimum diameter of the femur neck in a plane perpendicular to the head-neck midline" (1971:33). Because the focus of Van Gerven's research was not the sex assessment potential of individual measurements on the femur, he did not consider the MLD for sex assessment. However, his calculations for the MLD ($t = 7.90$, $P < 0.001$) indicated a statistically significant difference between males and females.

Knussman et al. (1993) conducted a study which addressed the sex assessment accuracy of the circumference of the femoral neck (CFN). Due to the oval shape of the femoral neck, a direct correlation of the CFN to the supero-inferior femoral neck diameter is impossible; however, one would expect that a sex-specific relationship would exist between the two. Based on the analysis of sex-specific means and standard deviations ($t = 9.6$, $P < 0.0005$), the CFN was shown to yield a statistically significant difference between the sexes. It may be assumed that the supero-inferior neck diameter would also yield a statistically significant difference. From the results of stepwise discriminant analyses, Knussman et al. concluded that neck circumference was the best predictor for sex in one analysis and the second best predictor in another.

Finally, the term "supero-inferior femoral neck diameter (SID)" used in the present study comes from two papers on australopithecine femoral fragments (Reed et al., 1993; Walker, 1973). Neither of the australopithecine studies used the SID as a method of sex assessment.

MATERIALS AND MEASUREMENT METHODS

The specimens used in this study were obtained from the Hamann-Todd collection and a variety of archaeological skeletal samples.

The Hamann-Todd collection

The Hamann-Todd skeletal collection, which is housed at the Cleveland Museum of Natural History, contains the defleshed remains of 3,157 autopsied individuals (531 females and 2,626 males) (Mensforth and Latimer, 1989). It constitutes the largest skeletal collection with accompanying demographic information in the world (Rothschild and Rothschild, 1995). The individuals in this collection, most of whom are African-Americans and Caucasians born between the years of 1825 and 1910, were assembled by T. Wingate Todd and other anatomists of Western Reserve University between 1910 and 1940. The skeletal remains were obtained from unclaimed cadavers following autopsy. These individuals were primarily of low socioeconomic status and were inhabitants of an early 20th century urban industrial community. The cadavers were subjected to defleshing using sodium hydroxide to remove the soft tissues (Rothschild and Woods, 1991). Additionally, the medical status and cause of death accompany most of the skeletal remains contained in the collection. Due to the nature of acquisition, the individuals represented in the Hamann-Todd skeletal collection are of known sex, age, and ancestry.

Archaeological collections

This study also utilized data from five archaeological skeletal samples to assess the differential preservation of the femoral neck in comparison to the femoral head. These data were collected from the National Museum of Natural History, Smithsonian Institution (Indian Knoll, Mobridge, Pachacamac, and Puyé), as well as the Department of Anthropology, Florida State University (Windover). Because all of these samples are of estimated sex, no attempt was made

to assess the sex prediction accuracy of the SID in archaeological samples. These samples were simply scored for the presence or absence of the femoral head (HVD) and femoral neck (SID) measurement areas; simple percentiles were calculated from the resulting data.

Element selection and measurement

There were a number of criteria by which the control sample specimens were selected. Because this was a study of the sex prediction capability of the SID in adult individuals, growth had to be complete, with reference to several indicators (fusion of epiphyseal plates, etc. (Bass, 1995)). The femoral neck at the point of measurement (the minimum diameter of the neck) did not show evidence of deterioration that could affect the accuracy of the measurement. There were also no signs of gross pathology, including severe cases of arthritis, rickets, congenital defects, or neck fractures on the specimens examined. The control sample of skeletal material from the Hamann-Todd collection totaled 203 individuals, including 52 African-American females, 51 African-American males, 50 Caucasian females, and 50 Caucasian males. The specimens cover an age range from 18 to 82 years with a mean age of 43.9 years (median = 41). A set of sliding calipers was used to span the neck of the femur at the minimum diameter in a supero-inferior direction (Fig. 1). Each specimen was measured three times. The calipers were reset following each measurement to assure accuracy and the average of the three measurements was recorded to the nearest hundredth of a millimeter.

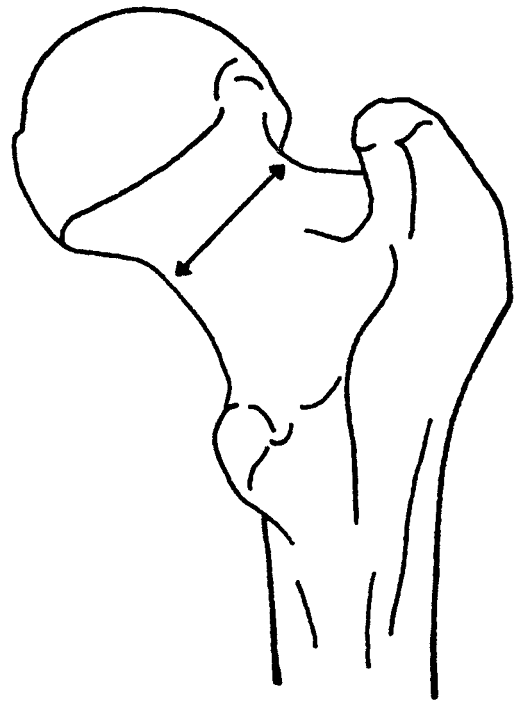

Fig. 1. The supero-inferior femoral neck diameter measurement (SID). The measurement is taken with sliding calipers across the narrowest part of the femoral neck.

STATISTICAL ANALYSES AND RESULTS

Analysis of the Hamann-Todd samples

The descriptive statistics for the Hamann-Todd sample are presented in Table 1. Statistics were calculated for the African-American and Caucasian samples (sexes combined), respectively, and for each sex and population group separately. The normality of each distribution was evaluated with the Wilk-Shapiro test. Wilk-Shapiro critical values and rankit plots (omitted here) were calculated for African-American males, African-American females, Caucasian males, and Caucasian females. Rankit plots were observed for positive linear relationships between observed SID values and n-scores. All four samples met this criteria. The observed correlations for African-American females (.983), African-American males (.984), Caucasian females (.983), and Caucasian males (.977) exceeded the Wilk-Shapiro critical value at the 5% level and, therefore, the normality assumption cannot be rejected for any group. Although normality cannot be proven, it appears that any deviations will have little effect on the outcome of standard parametric methods.

Assessment of sex and population effects on SID. Differences in femoral morphology between Caucasians and African-Americans have been observed previously (Craig, 1994; Stewart, 1962) and, as such, verification of similar SID measurements for both groups is necessary. A two tailed *t*-test (α =

TABLE 1. SID (mm) descriptive statistics for sexed individuals in the Hamann-Todd collection

Statistic	Caucasian combined	African-American combined	Caucasian males	Caucasian females	African-American males	African-American females
N	100	103	50	50	51	52
Minimum	24.59	23.74	29.40	24.59	29.29	23.74
Maximum	39.62	36.74	39.62	31.85	36.74	31.60
Median	30.57	29.48	32.86	28.12	31.77	26.98
Mean	30.69	29.59	33.53	27.86	31.93	27.31
SD*	3.45	2.87	2.20	1.67	1.73	1.66
Variance	11.88	8.23	4.83	2.78	3.01	2.73

* Standard deviation.

TABLE 2. ANOVA for population and sex-specific SID measurements

Source	Degrees of freedom	Sum of squares	Mean square	F-statistic	*P*-value
Sex	1	1341.61	1341.61	402.00	.0001
Population	1	58.22	58.22	17.45	.0001
Sex × Population	1	13.99	13.99	4.19	.042
Error	199	664.11	3.34		

Bonferroni multiple comparison procedure.
Family error rate = .0412, Individual error rate = .0083, critical value = 2.665.

0.05) rejected the null hypothesis of equal Caucasian and African-American SID measurements (df = 192.5, $P = 0.015$). Satterthwaite and pooled variance tests produced identical *P*-values. As indicated in Table 1, Caucasians exhibit larger femoral neck diameters than do African-Americans. It seems prudent to perform statistical analyses for each population sample separately.

We further evaluated sex and population-specific differences in sexual dimorphism using an ANOVA supplemented with Bonferroni alpha protection. The ANOVA model is presented in Equation 1 and the results are presented in Table 2.

$$\text{SID} = \text{Sex} + \text{Population} + \text{Sex} * \text{Population} + \text{error} \tag{1}$$

The individual effects are both significant at the $P = 0.0001$ level, indicating that both sexual and population differences are exhibited in the SID measurement. The *P*-value for the interaction effect indicates a statistically significant F-statistic for the model as a whole. The ANOVA model explains 68% of the variability in the femoral neck diameter for the study sample. No significant differences exist between Caucasian and African-American female samples. The African-American male sample was statistically greater than both female groups. The Caucasian male sample was statistically greater than the African-American male and female samples as well as the Caucasian female sample. In other words, the female samples were grouped together, with each male sample grouped separately. The Bonferroni family error rate (FER) indicates a 4.1% chance of claiming a statistical difference when in fact none exists (a Type I error). Additional runs using Fisher's Least Significant Difference method (FER = .202) and Tukey's Honest Significant Difference method (FER = .05) produced identical groupings of the samples. The Bonferroni method is preferred and reported here due to its conservatism and low family error rate.

The ability of the SID measurement to accurately predict sex was evaluated using the Systat for Windows version 6.0 statistical package. The data were analyzed using a linear discriminant function analysis, which produced a single discriminant function with a breakpoint set equal to zero. Scores above zero are males and scores below zero are females. A random sample of 75 African-Americans and 75 Caucasians were selected to develop the predictive model.

The discriminant function for African-Americans is:

$$\text{Sex} = 0.578 * \text{SID} - 17.141; \tag{2}$$

where Sex < 0 indicates a female and Sex > 0 indicates a male The discriminant function for Caucasians is:

$$\text{Sex} = 0.496 * \text{SID} - 15.163; \tag{3}$$

where Sex < 0 indicates a female and Sex > 0 indicates a male.

The Systat package also produced classification matrices for each linear discriminant

TABLE 3. Classification matrices for African-American sample

	Normal matrix		
Estimated	Male	Female	% Correct
Actual			
Male	45	6	88
Female	4	48	92
Total	49	54	90
	Jackknifed matrix		
Estimated	Male	Female	% Correct
Actual			
Male	45	6	88
Female	4	48	92
Total	49	54	90

TABLE 4. Classification matrices for Caucasian sample

	Normal matrix		
Estimated	Male	Female	% Correct
Actual			
Male	45	5	90
Female	3	47	94
Total	48	52	92
	Jackknifed matrix		
Estimated	Male	Female	% Correct
Actual			
Male	45	5	90
Female	3	47	94
Total	48	52	92

TABLE 5. Classification matrices for pooled African-American and Caucasian samples

	Normal matrix		
Estimated	Male	Female	% Correct
Actual			
Male	94	7	93
Female	8	94	92
Total	102	101	93
	Jackknifed matrix		
Estimated	Male	Female	% Correct
Actual			
Male	94	7	93
Female	8	94	92
Total	102	101	93

function. The normal and jackknifed matrices for African-Americans and Caucasians are presented in Tables 3 and 4, respectively. For African-Americans, 65 of 75 individuals (87%) were sexed correctly using Equation 2. The jackknifing procedure produced identical results. An independent, random holdout sample consisting of 15 females and 13 males were sexed using Equation 2. Twenty-eight of 28 individuals were correctly classified. For Caucasians, 69 of 75 individuals (92%) were sexed correctly using Equation 3. Again, the jackknifing procedure produced identical results. An independent, random holdout sample consisting of 13 males and 12 females were sexed using Equation 3. Twenty-four of 25 individuals were correctly classified.

Because it is somewhat contradictory to develop a method of sex assessment for fragmentary remains and simultaneously require that the population from which the individual was drawn be known, a linear discriminant function analysis was calculated in the same manner as above for the pooled African-American and Caucasian (AAC) sample. A total of 150 individuals were used to generate the discriminant function. The equation for individuals of unknown ancestry is:

$$\text{(pooled AAC): Sex} = 0.510 * \text{SID} - 15.356; \tag{4}$$

where Sex $<$ 0 indicates a female and Sex $>$ 0 indicates a male.

The classification matrices (Table 5) indicate that 135 of 150 individuals (90%) were sexed correctly using Equation 4. The jackknifing procedure again produced identical results. To further evaluate the accuracy of the predictive model, we classified an independent, random holdout sample consisting of 53 individuals. The holdout sample contained both males and females of African-American and Caucasian ancestry. All 53 individuals in this holdout sample were correctly classified using the discriminant function presented above. We had anticipated that the holdout samples would provide a more accurate assessment of the true ability of this predictive model to predict sex. It is clear, however, that the accuracy for the holdout sample is an overestimation. It is likely that the true predictive accuracy in the population is closer to 90% as evidenced by the classification matrices in Tables 3, 4, and 5.

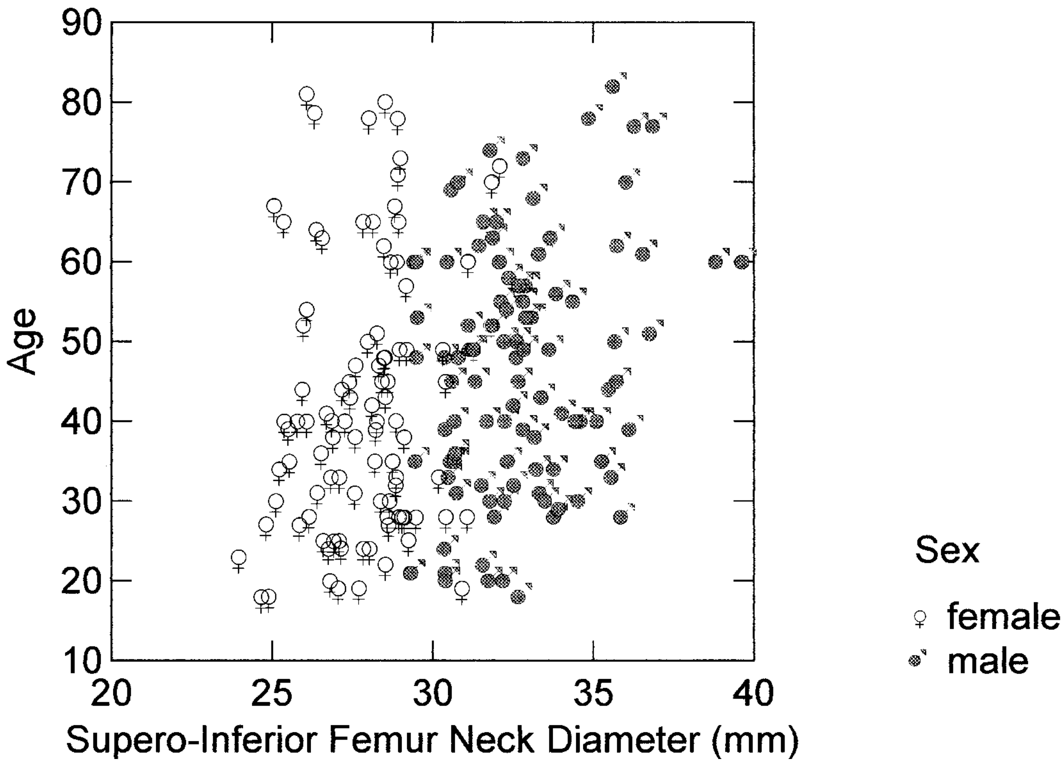

Fig. 2. This graph is a scatter plot of SID measurements by age. The data presented in this graph, which includes all 203 observed individuals in the Hamann-Todd collection, indicates a clear separation between male and female SID measurements; however, some overlap does occur. This graph also presents evidence in support of the contention that the SID measurement remains stable with regard to aging.

Assessments of age changes in the femoral neck

In looking at the age distribution of the SID measurements in Figure 2, there do not appear to be significant differences in the measurement according to age. The Pearson correlation coefficient between age-at-death and the SID equals .244. Age-at-death explains 6% of the variability in the SID measurements. We further evaluated age-dependent changes in the SID using an ANCOVA model with age-at-death as the continuous covariate. The ANCOVA model is presented in Equation 5 and the results of the analysis are presented in Table 6.

$$\text{SID} = \text{Sex} + \text{Population} + \text{Sex} * \text{Population} + \text{Age} + \text{error} \tag{5}$$

Where sex, population, and the sex/population interaction effects produced signifi cant F-statistics, age-at-death was not a significant component in the model ($P = 0.088$).

TABLE 6. ANCOVA testing for age-related changes in femoral neck diameter

Source	Degrees of freedom	Sum of squares	Mean square	F-statistic	*P*-value
Sex	1	1278.52	1278.52	386.83	.0001
Population	1	29.72	29.72	8.99	.003
Sex × Population	1	13.81	13.81	4.18	.042
Age	1	9.70	9.70	2.94	.088
Error	198	654.41	3.31		

Preservation of the femoral neck in archaeological samples

A total of 424 femoral elements were randomly selected from the collection facilities at the Smithsonian Institution and Florida State University. Each femur was scored for the integrity of the femoral head

TABLE 7. Preservation rates of SID vs. HVD in archaeological samples

Site	Windover	Mobridge	Pachacamac	Puyé	Indian Knoll	Totals
n	169	23	134	58	40	424
Head present (%)	50.3	47.8	38.1	39.7	90.0	48.6
Neck present (%)	76.9	78.3	100.0	72.4	100.0	85.8
No measurable feature (%)	16.6	13.0	0.0	27.6	0.0	11.1

and SID measurements. The results are presented in Table 7. Of the 424 femora observed, only 48.6% had a measurable femoral head diameter; the femoral neck was measurable in 85.8% of the elements. Only 11.1% were so poorly preserved that no sex demographic data could be obtained from the element. Differences per sample range from 10 to 61.9%.

DISCUSSION

The statistical evidence indicates that the SID is a reliable sex predictor for American skeletal samples. Percent accuracies range from 87% (for African-Americans) to 92% (for Caucasians) in this study sample. These prediction rates are comparable to published data for sex estimation using the femoral head diameter (Dwight, 1904; Stewart, 1979; Bass, 1995). However, one of the primary concerns of the authors was the actual applicability of the SID measurement. A feature with an extremely high sex prediction accuracy that rarely preserves well enough is not very useful. This is particularly true in extremely fragmentary samples where the preferred multivariate sexing techniques are often unavailable due to incomplete preservation. In terms of preservation, this study is concerned with the proximal end of the femur, and if encountered in a sample, the likelihood of the upper third of the bone providing sex demographic data.

Although taphonomic processes are applicable to both forensic and archaeological samples, it was impossible to examine preservation-related issues in the Hamann-Todd collection. Therefore, preservation data were collected for five archaeological sites (Windover, Florida; Mobridge, South Dakota; Pachacamac, Peru; Puyé, New Mexico; and Indian Knoll, Kentucky) to provide a range of environmental settings from which to derive the preservation data.

These data suggest that the femoral neck exhibits a much higher rate of intact preservation than the femoral head. If one considers differential preservation of these features for each sample, the femoral neck is consistently present almost twice as often as the femoral head.

A word of caution is necessary when considering these data. The Pachacamac population was curated differently than the other four samples, which may have affected the preservation rate for this site. Where Windover, Mobridge, Puyé, and Indian Knoll were curated according to discrete individuals, the Pachacamac sample was curated by element. Essentially, the authors pulled a drawer filled only with left femora. It was observed that the deeper the femora were located in the drawer, the poorer the condition of the femoral heads became. It was apparent that some of the deterioration was the result of post-excavation handling. This would account for the extreme variability between the number of preserved necks versus heads in the Pachacamac sample. These data suggest that different curation methods can exert a significant effect on problems of preservation; the mere anatomical position of the femoral head appears to compromise its utility.

Recent work in paleodemography has also demonstrated that age-related changes in skeletal morphology can introduce sexing biases in skeletal analysis (Walker, 1995). In light of this fact, the authors felt it necessary to test for age-related differences in the femoral neck diameter. Ideally, a longitudinal study would provide the most accurate assessment of age-related changes in femoral neck morphology. However, such a study is not presently available.

Our analysis of what is available, African-American and Caucasian samples, showed no bias due to obvious age effects. Indeed, the fairly weak correlation between age-at-death and the SID, combined with the insig-

nificant *P*-value for the age covariate, indicates that age effects are minimal in this methodology.

These conclusions are given further support by Beck et al. (1992), who conducted a study on the femoral neck that addressed apparent strength differences between males and females. This study, conducted on 409 individuals between the ages of 19 and 93, examined sex-specific differences in the restructuring of the femoral neck with age. Although the analyses performed in this study considered strength and geometric structuring, which are not directly related to the present study, the authors of the 1992 study present data further substantiating the stability of the femoral neck when examined in relation to age. Beck et al. examined age as a factor for change on the femoral neck width (a measure comparable to the SID). The authors of the 1992 study concluded that aging has no effect on femoral neck width. Beck et al. (1992) provided a plot of the femoral neck width versus age which represents the same trends as Figure 2 of the present study.

SUMMARY AND CONCLUSIONS

The femoral neck diameter may provide an accurate means of sex assessment when the more common multivariate sexing criteria and the established univariate postcranial features are unavailable due to incomplete preservation. The study sample for this analysis consisted of African-American and Caucasian individuals obtained from the Hamann-Todd collection. Linear discriminant functions were developed for both population-specific samples and for the combined African-American and Caucasian (AAC) sample. Jackknifed classification matrices produced error rates in the 80 to 90% range. Independent validation using holdout samples produced unrealistically high accuracies ranging from 96% for the Caucasian sample to 100% for the African-American and combined AAC samples. True error rates are likely much higher. We are unable to comment at this time on the applicability of this methodology to archaeological or modern skeletal populations. However, the high rate of intact preservation of this feature makes future work on this anatomical landmark promising.

ACKNOWLEDGMENTS

The authors thank Dr. Michael F. and Mrs. Diane M. Seidemann for providing funding for the initial research trip to Cleveland. The authors thank Lyman Jellema and Dr. Bruce Latimer of the Cleveland Museum of Natural History. Thanks to Drs. Hunt, Owsley, and Ubelaker at the National Museum of Natural History, Smithsonian Institution. Thanks also to James Halpern for the artwork (Fig. 1) and Dr. David N. Dickel for his assistance during the 1996 data collection and for his advice regarding the present study. The statistical advice of Drs. Joseph Powell and Steven Nawrocki, as well as editorial comments by Ericka Sondergaard, were invaluable. Thanks also to Melinda and Elizabeth Muth for their support throughout this project.

LITERATURE CITED

Bass WM (1995) Human Osteology: A Laboratory and Field Manual, Fourth Edition. Columbia, MO: Missouri Archaeological Society.

Beck JB, Ruff CB, Scott WW, Plato CC, Tobin JD, and Quan CA (1992) Sex differences in geometry of the femoral neck with aging: A structural analysis of bone mineral data. Calcif. Tissue Int. *50:*24–29.

Craig EA (1994) Bones of the Knee Joint and Individual Features That can be Used for Forensic Identification. Ph.D. dissertation. University of Tennessee, Knoxville.

Dwight T (1904) The size of the articular surfaces of the long bones as characteristic of sex: An anthropological study. Am. J. Anat. *4:*19–31.

Knussmann R, Van Vark GN, Hartman BR, and Den Arend A (1993) On the question of size-independent sex-specific differences in human long bones. Homo *44(3):*242–262.

Mensforth RP and Latimer BM (1989) Hamann-Todd collection aging studies: Osteoporosis fracture syndrome. Am. J. Phys. Anthropol. *80:*461–479.

Paterson RS (1929) Radiological investigation of epiphyses of long bones. J. Anat. *64:*28–46.

Reed KE, Kitching JW, Grine FE, Jungers WL, and Sokoloff L (1993) Proximal femur of *Australopithecus africanus* from Member 4, Makapansgat, South Africa. Am. J. Phys. Anthropol. *92:*1–15.

Rothschild BM and Rothschild C (1995) Comparison of radiologic and gross examination for detection of cancer in defleshed skeletons. Am. J. Phys. Anthropol. *96:*357–363.

Rothschild BM and Woods RJ (1991) Spondyloarthropathy: Erosive arthritis in representative defleshed bones. Am. J. Phys. Anthropol. *85:*125–134.

Stewart TD (1962) Anterior femoral curvature: Its utility for race identification. Hum. Biol. *34:*9–62.
Stewart TD (1979) Essentials of Forensic Anthropology. Springfield, IL: Thomas.
Van Gerven DP (1971) The Contribution of Size and Shape Variation to Patterns of Sexual Dimorphism of the Human Femur. Ph.D. dissertation, University of Massachusetts.
Van Gerven DP (1972) The contribution of size and shape variation to patterns of sexual dimorphism of the human femur. Am. J. Phys. Anthropol. *37:*49–60.
Walker A (1973) New *Australopithecus* femora from East Rudolf, Kenya. J. Hum. Evol. *2:*545–555.
Walker PL (1995) Problems of preservation and sexism is sexing: Some lessons from historical collections for paleodemographers. In: SR Saunders and A Herring (eds.): Grave Reflections, Portraying the Past Through Cemetery Studies. Toronto: Canadian Scholars Press, pp. 31–47.
Walmsley T (1915) Observations on certain structural details of the neck of the femur. J. Anat. Physiol. *49:*305–313.